DTU Library

Data-driven engineering design research: Opportunities using open data

Parraguez Ruiz, Pedro; Maier, Anja

Published in:
Proceedings of the 21st International Conference on Engineering Design (ICED17), Vol. 7: Design Theory and Research Methodology

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Parraguez Ruiz, P., & Maier, A. (2017). Data-driven engineering design research: Opportunities using open data. In *Proceedings of the 21st International Conference on Engineering Design (ICED17), Vol. 7: Design Theory and Research Methodology* (pp. 41-51). Design Society.

DATA-DRIVEN ENGINEERING DESIGN RESEARCH: OPPORTUNITIES USING OPEN DATA

Parraguez, Pedro; Maier, Anja
Technical University of Denmark, Denmark

Abstract

Engineering Design research relies on quantitative and qualitative data to describe design-related phenomena and prescribe improvements for design practice. Given data availability, privacy requirements and other constraints, most empirical data used in Engineering Design research can be described as “closed”. Keeping such data closed is in many cases necessary and justifiable. However, this closedness also hinders replicability, and thus, may limit our possibilities to test the validity and reliability of research results in the field. This paper discusses implications and applications of using the already available and continuously growing body of open data sources to create opportunities for research in Engineering Design. Insights are illustrated by an examination of two examples: a study of open source software repositories and an analysis of open business registries in the cleantech industry. We conclude with a discussion about the limitations, challenges and risks of using open data in Engineering Design research and practice.

Keywords: Information management, Research methodologies and methods, Open source design, Data-driven design, Open data

Contact:
Dr. Pedro Parraguez Ruiz
Technical University of Denmark
Department of Management Engineering
Denmark
ppru@dtu.dk

Please cite this paper as:
Surnames, Initials: *Title of paper*. In: Proceedings of the 21st International Conference on Engineering Design (ICED17), Vol. 7: Design Theory and Research Methodology, Vancouver, Canada, 21.-25.08.2017.

1 INTRODUCTION

The development and growth of scientific fields depends on their capability to collectively acquire, analyse and interpret data (Nielsen, 2011; OECD, 2015). In recent years, advancements in information technologies, data science, and instrumentation have dramatically increased the stream of data and the analytical capabilities available for research (Hey et al., 2009). This increase of data and new analytical capabilities has also substantially expanded the frontiers of what is possible in industrial practice (Provost and Fawcett, 2013). As a result, what has been described as a data-driven revolution has reshaped both industrial and scientific landscapes, shifting resources, attention and citations (Nielsen, 2011; Nosek et al., 2015).

An important element and enabler of this data revolution is the growing use of and demand for **open data,** which has led to an increase in transparency, accountability and which has allowed for more research replicability and re-use. All of which ultimately contributes to improving the reliability and validity of research results (Pampel and Dallmeier-Tiessen, 2014). An example of this growing demand and support for open data are the *Transparency and Openness Promotion (TOP)* guidelines for journals (Nosek et al., 2015).

In this position paper, we propose that it is important and timely to review the opportunities of using open data in Engineering Design research, analysing its implications and applications to strengthen design science and practice. For this purpose, the remainder of the paper is structured as follows: Section 2 examines data related topics in Engineering Design, in particular those related to the use of closed data and the opportunities of using open data. Section 3 provides a review of implications and applications of using open data in published Engineering Design research. Section 4 further elaborates two concrete application cases and section 5 closes with a discussion and conclusions.

2 DATA AND ENGINEERING DESIGN RESEARCH

Following the data revolution experienced in the life sciences, physics and other social and natural sciences (Hanson et al., 2011; Hey et al., 2009; Nielsen, 2011), Engineering Design research also shows signs of becoming more data-intensive, further developing its empirical grounding and incorporation of multiple types of data and analysis methods into research methodologies (Blessing and Chakrabarti, 2009; Cash et al., 2016). For instance, Engineering Design research acquires quantitative and qualitative data in both laboratory and industrial environments using methods that include observations, questionnaires, interviews, and experiments. In addition to observation and experimentation, empirical data is also acquired through other structured and unstructured data sources, including archival data such as documents, log-files or sensor data obtained from companies; a type of data that is likely to increase due to the extensive digitalisation of the engineering design process. As an example of the increased predominance of empirical data, our analysis of close to 7.000 scientific articles from the Engineering Design field reveals that the proportion of articles mentioning in their title, keywords or abstract a term related to the use of empirical data has steadily increased, growing from less than a third in 1979 to more than 50% in 2016 (see Figure 1).

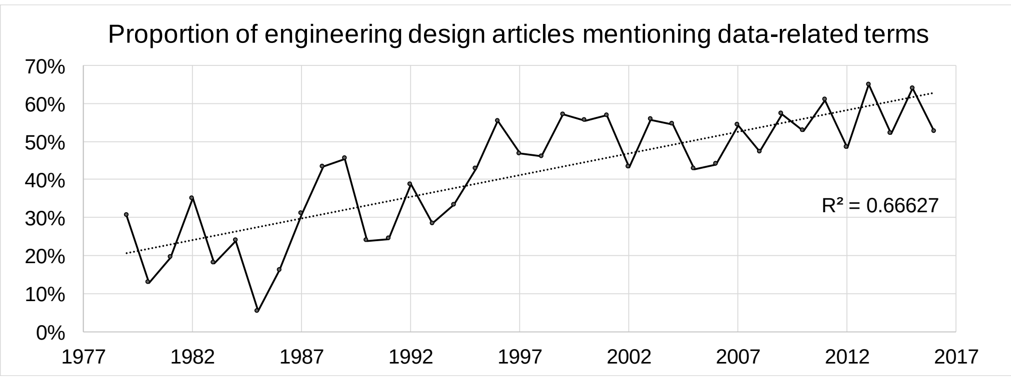

Figure 1. Proportion of engineering design articles mentioning data-related terms (1979-2016). The figure includes all articles for the seven journals listed on the Design Society website and all papers indexed for DESIGN and ICED. Accessed on 05/Nov/2016 Source available as open data at: https://doi.org/10.5281/zenodo.204820

Due to privacy requirements and other constraints related to the private, sensitive and/or proprietary nature of the acquired data, most empirical data used in Engineering Design research remains closed. Given the previous background, in what follows we examine why the use of open data is worth being further explored and exploited in Engineering Design research and how open data can be leveraged to a) increase the overall replicability of Engineering Design research, allowing revisiting previous research questions, hypothesis and assumptions and b) create and test new research questions and hypotheses. Both uses that are expected to benefit the overall scientific development of the field and design practice.

2.1 Scope and operational definition for open data in Engineering Design

This paper focuses on applications and implications of using already available open data sources to conduct Engineering Design research. Although this topic is connected and complementary to themes such as open science, scientific data management, open access, big data and research on open design, these topics, while important, are outside of our scope.

In the context of this paper, the attribute of "open" refers to the original condition of the data source. In other words, it refers to the openness that directly affects the data acquisition stage, and not data preparation, analysis, interpretation or final dissemination. More specifically, and following the Open Knowledge International (2012) definition, open data can be understood as: *"Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike."*

However, in practice, most data sources are not completely open or closed. Instead, data usually fall within a spectrum of different levels of requirements and restrictions. Consequently, in this paper data will be considered open for Engineering Design research purposes if:

- It is already in the public domain, or there is a transparent, non-exclusive and publicly-posted procedure to request the data.
- Data permissions and restrictions, if they exist, allow using the data in a research-context.
- The data is already available in digital form, as this significantly reduces acquisition barriers.

2.2 Data-related challenges in Engineering Design research relevant to open data usage

Data is a common discussion topic in Engineering Design research, especially in the context of design research methodology and methods. Moreover, data-related discussions often appear associated, directly or indirectly, with the challenges and barriers of acquiring and using "closed" data within research practice. In what follows, we list a selection of challenges and barriers stated in previous Engineering Design literature that are associated with the use of closed data as well as with opportunities related to using open data:

- "More and more researchers point out the need to go further and set up a comprehensive, archival database of research results to allow these to be compared and integrated…" "… Bringing the results together is one of the pre-requisites for the development of comprehensive models and theories." (Blessing and Chakrabarti, 2009, p. 7)
- "…data collection should focus on data that can be collected within the constraints of the research project. If the latter is not the case, the research questions and hypotheses need to be adapted." (Blessing and Chakrabarti, 2009, p. 89)
- "…it is through rigour and standard approaches that the body of knowledge can increase and be linked together and raw data can be shared. This will give the results of design research real credibility and real traction to practitioners of all sorts." (Cash et al., 2016, pt. Foreword)
- "The capability to harness and leverage data from non-traditional sources is crucial. Open data efforts, which do not make use of sensitive or protected data, often do not require much political clout or resources to begin, i.e., projects can be initiated without first obtaining access to restricted data." (Chiarini et al., 2014, p. 76)
- "…opening access to knowledge (in an open data manner) that is created during the open design process is becoming more common. Research has yet to address how the data commons, which is produced in the design process, can be useful, for whom it can be useful, and in which format it should be published." (Aitamurto et al., 2013)

Finally, an EPSRC report following a workshop on *Engineering Visions in Design* documented in Gillespie (2010), emphasised the role of data and its future in Engineering Design research and practice. Some highlights of data-related discussions included the great potential of data-driven models to better understand behaviour and uncertainty in design, as well as the relevance of embracing new data sources, open data directives, and data methods to tap into the wealth of digital data about natural events and human activities. Some of the main drivers mentioned in the workshop included the need to use shareable datasets so that a) design research can be validated, b) built upon and c) the cost of data acquisition can be lowered.

One may summarise the quotes selected above by highlighting that closed data limits the replicability of the research process and results, making it harder to judge the validity and reliability of research results. **Closed data also prevents** researchers from **linking, validating, enriching, reusing and re-sharing data sources, hindering** the possibility of building on each other's work. While **open data** have been identified by some researchers as an opportunity to further develop Engineering Design research, this paper argues that we need to develop the opportunities further and need to consider specific implications and applications.

We take these challenges and opportunities as a starting point to explore implications and applications of open data usage in the Engineering Design field in more detail through concrete examples. Our objective is to contribute to the specifics of how Engineering Design research may be equipped to better leverage open data sources.

3 OPEN DATA IN ENGINEERING DESIGN RESEARCH: IMPLICATIONS AND APPLICATIONS

In this section, we present implications and concrete applications for the use of open data in Engineering Design research. Although the focus is on the Engineering Design field, we will also draw on the experience of other fields to identify insights valuable for the future development of Engineering Design research.

3.1 Implications

In a survey conducted by the journal Science, to which about 1.700 of its peer-reviewers responded to (Hanson et al., 2011), more than 44% of the surveyed scientists replied that they "often" or "half of the time" use datasets from the published literature or archival databases (open data) for their original research papers. The rest responded that they use these sources, but only rarely. Considering the growth of open science since then, and the wider availability of open data repositories, it is likely that today, these figures are higher. Although we have no comparable figures for Engineering Design research, the challenges listed in the previous section and our analysis of data usage in Engineering Design suggest that these figures are probably lower within the Engineering Design community. Considering that the use of open data is positively associated with the strengthening of scientific disciplines, higher levels of innovation and overall societal impact (OECD, 2015; Pampel and Dallmeier-Tiessen, 2014), it seems relevant to identify opportunities to increase open data usage in Engineering Design research.

Whilst the use of open data might be more pervasive in other scientific disciplines, there are also still important data-related challenges affecting life- and social sciences across the board. These general challenges originate in regulatory, technical and social issues which also apply to Engineering Design Research. For example, data protection and privacy issues are common reasons why data sharing might not be feasible or limited (Pampel and Dallmeier-Tiessen, 2014). Moreover, in a study conducted by Tenopir et al. (2011) with more than 1.300 researchers, 67% claimed a ''lack of access to data generated by other researchers or institutions'', but at the same time only about a third (36%) of the respondents agreed that "others can access their data easily''. This indicates that scientist appreciate the importance and value of accessing open research data. At the same time this also shows that scientists are far less inclined to take the steps required to increase the accessibility of their own data. Such asymmetry is, among other things, the result of misaligned incentives and insufficient support to data sharing activities (Bartling and Friesike, 2014).

In addition to the challenges and implications related to the low usage and availability of open data in Engineering Design, the intrinsic characteristics of many open data sources bring their own challenges and implications to Engineering Design research. For example, open data is increasingly also "big data", in the sense that its volume (i.e. amount of data), variety (i.e. combination of different types of structured and unstructured data) and velocity (i.e. speed at which is generated) are large enough to represent challenges in their own right. Additionally, open data is usually gathered for a different purpose than what the researcher originally envisaged. This means that researchers using open data have to rely on a third party for aspects such as data veracity and quality.

Furthermore, implications of using open data extend to all the stages of the research process, including the stages of data acquisition, data analysis and data interpretation (Pampel and Dallmeier-Tiessen, 2014). Such wide implications require the acquisition of a set of data-intensive interdisciplinary skills and resources (Provost and Fawcett, 2013) which are not traditionally found in Engineering Design research. For instance, for **data acquisition**, open data sometimes has to be gathered through methods such as web crawling and other data harvesting techniques, and subsequently normalised and structured in suitable databases. For **data analysis**, traditional statistical analysis often have to be combined with complex exploratory analysis, content analysis, machine learning and other techniques often included under the umbrella of the growing field of data science. Finally, for **data interpretation**, a solid understanding of the results of the analysis and their limitations greatly benefits from interactive visualisations that allow for the exploration of complex patterns, which are hard to summarise in static tables and charts. Thus, to fully leverage the value of open data sources for Engineering Design research, it is not only necessary to acknowledge its potential usefulness, identify relevant open data sources and applications, but it is also necessary to address the regulatory, technical and social challenges previously mentioned, including investing in new skills and data-related capabilities.

3.2 Applications

Having in section 3.1 described implications of using open data in Engineering Design research, in this section we provide concrete application examples and list the main types of open data sources that are relevant and already available to be applied in an Engineering Design research context. Our results are presented in Table 1, where we include for each type of data: its potential uses for engineering design, online sources where that type of data can be harvested, and relevant references that provide examples of the usage of each data type in the context of Engineering Design research. We generated this list mining the almost 7.000 research papers of the Engineering Design field for examples of open data usage (the same data used for the analysis of data-related terms in section 2 for data-related terms), surveying and curating open data repositories such as the ones provided by the Global Open Data Index and other Data Science platforms (e.g. http://index.okfn.org/dataset and www.kaggle.com), and combined this with the authors own experience using open data.

The results summarised in Table 1 show the relatively recent emergence of articles using open data sources in research related to Engineering Design subjects (see publication year). They also show several authors repeated within or across source types as well as types of open data where we had to look outside the core Engineering Design journals and conferences to find relevant examples. All this suggests that we might be currently observing some of the early stages of open data usage in our field, and more research examples will over time build up with the growth of new data-related skills and capabilities.

Table 1. Types of open data sources identified as relevant for Engineering Design research

Open data sources	Data source examples	Relevance for Engineering Design Research (examples)	Research examples in Engineering Design
Patents Repositories including full patent text and related metadata.	www.uspto.gov www.epo.org https://patentscope.wipo.int	-Evolution of technical systems -Macro connections between technical functions and behaviour (using design description and patent claims)	(Kang and Tucker, 2015; Luo et al., 2014; Song, Triulzi, et al., 2016; Song, Yoon, et al., 2016)
Scientific Publications Repositories include scientific articles and their metadata.	https://scholar.google.com www.worldcat.org https://citeseerx.ist.psu.edu www.archive.org www.scopus.com	-Meta research on engineering design -Trends and evolution of technical artefacts	(Chai and Xiao, 2012; Hamraz et al., 2013; Parraguez and Maier, 2012a)
Search History Anonymised public registries of queries in search engines	www.google.com/trends http://googletrends.github.io/data	-Evolving patterns in technical requirements as expressed by consumers in their searches	(Spee and Basaiawmoit, 2016; Tucker and Kim, 2011)
Knowledge Databases Open repositories of information created by organisations and individuals including public wikis.	www.wikipedia.org http://www.appropedia.org www.wikidata.org https://concept.research.microsoft.com	-Creation of bottom up knowledge taxonomies to aid engineering design research -Useful for triangulation against knowledge acquired through other sources	(Albers et al., 2010; Walthall et al., 2011)
Business Registries and Directories Databases listing companies and their metadata including industrial sector, location, size, etc.	https://datacvr.virk.dk www.handelsregister.de http://index.okfn.org/dataset/companies https://opencorporates.com	-Mapping of knowledge, organisational and technology landscapes -Analysis of industrial and technological evolution	(Michelino et al., 2015; Parraguez and Maier, 2012b, 2016)
Company Websites Public websites that can be mined to harvest structured and unstructured data.	Any open company website	-Information about industries, products, needs, etc.	(Fichter, 2009; Staudenmayer et al., 2005; Unger and Eppinger, 2009; Wang et al., 2013; Zhou, 2013)
Product Forums and other Discussion Lists Open online forums for technical reports, complaints, suggestions, troubleshooting, reviews, etc.	http://stackoverflow.com www.stackexchange.com www.amazon.com	-Analysis of functionality, features, quality perception and other variables from a bottom-up perspective.	(Raghupathi et al., 2014; Storga et al., 2013; Tucker and Kim, 2011; Wang et al., 2016)
Social Networks Including large social media networks such as Twitter, but also niche social networks for developers and engineers.	https://twitter.com Public posts on Facebook.com and Linkedin.com Social components of sites such as Github.com	-Mining interactions between engineers, engineers and customers, information spread, etc.	(Pajo et al., 2015; Stone and Choi, 2013; Tuarob and Tucker, 2015)
Open innovation and open challenge platforms Platforms that publicly distribute industry challenges to, for example, crowdsource their solutions	www.innocentive.com www.yet2.com www.ninesigma.com www.openideo.com	-Macro analyses of technology needs and trends. -Descriptive and prescriptive studies related to the match of technology pull and push.	(Pourmohamadi and Dong, 2011)
Repositories of open design results and open source software Online sources where is possible to download designs of artefacts and software code.	www.thingiverse.com https://github.com https://grabcad.com http://opensourceecology.org	-Analysis of the evolution of engineering design objects over time including aspects of distributed team work and design variants.	(Bianchi et al., 2015; Le and Panchal, 2012; MacCormack et al., 2006)
Social, economic and population macro-data resources Including industry figures, population trends, etc.	http://data.worldbank.org https://data.oecd.org https://datausa.io http://ec.europa.eu/eurostat	-Analyse the influence of macro exogenous factors on individual engineering design projects or firms.	(Hidalgo et al., 2007; Luo et al., 2014; Makumbe et al., 2009)

4 ILLUSTRATION OF TWO OPEN-DATA APPLICATION CASES

In what follows, we offer an examination of two of the examples presented in table 1 to illustrate the use of open data in the context of design research studies: a) the **use of open source software repositories** performed by Le and Panchal (2012) and b) the use of open **business registries** performed by Parraguez and Maier (2016)

Le and Panchal (2012) use Drupal's open online repository (an open-source software product) to study the mutual influence of product and community architectures. The examined repository includes information for both product and community structures for multiple versions of the software, has data on 7000 community-contributed add-ons and registers, and registers the actions and interactions of more than 5000 developers working on each version and add-on. As such, this repository provides a wealth of data about large product development processes, which Le and Panchal (2012) use to analyse the interdependent co-evolution of product structures and community structures using dependency modelling techniques. Based on Le and Panchal's analysis, the results are that, in the case of Drupal, product structures significantly influence community structures, testing what is frequently described as the "mirroring hypothesis". Here open data sources allowed tackling a challenging research question, leveraging a rich data source which is already in the public domain.

In a research project called "Net-Sights" Parraguez and Maier (2016) use a combination of open business registries and directories of the Danish cleantech industry to: 1) study and predict potential new product development collaborations and 2) provide a descriptive map of interlocking technologies in the Danish cleantech industry. Their open data sources include State of Green's public directory (www.stateofgreen.com) and the Central Business Registry of Denmark (CVR) (https://datacvr.virk.dk). State of Green provides a curated list of over 600 cleantech organisations with information on the company product and services, industrial sector and sub-sectors, geographical location, and previous collaborations in engineering projects. The CVR is the official registry of companies in Denmark, providing access to information from over 700.000 companies with metadata that covers number of employees, basic financial information, data of foundation, industry branch, etc. These two open data sources were combined with factors reported in previous Engineering Design literature to provide a new way of mapping the technological capabilities and collaboration opportunities within the Danish cleantech industry. The results of Net-Sights include a conceptual approach grounded in Engineering Design research, a data-driven web application (www.netsights.dk) developed to visually explore and identify collaborative potential between companies (see Figure 2 below), and a platform that can export the results as open data, so that the models and analysis can be re-used by researchers and practitioners.

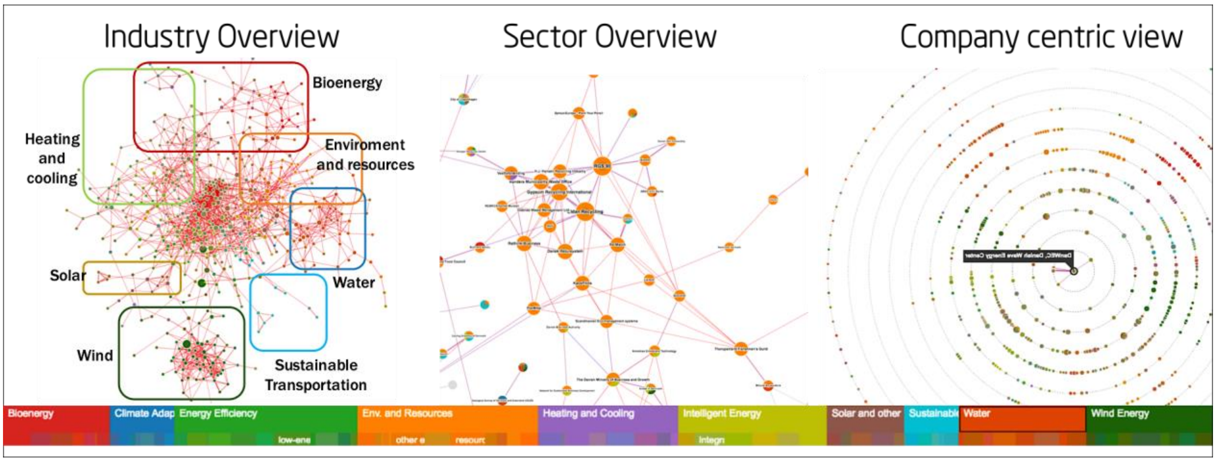

Figure 2. The Danish Cleantech landscape showing cpmpanies and industry sub-sectors as visual output from the web application usable for research and decision-making. Source: www.netsights.dk

These cases show two examples that illustrate the type of data available, its scale, and the relevance of these open data sources to tackle old and new research questions in Engineering Design. These examples also provide a glimpse into the new challenges and requirements related to data acquisition, analysis and interpretation that these types of research projects face.

5 DISCUSSION AND CONCLUSION

This paper has presented an overview of the implications and concrete applications of using open data in Engineering Design research, including specific data-sources and their possible uses in Engineering Design research and practice (Table 1). We have argued that in order to strengthen the scientific development of Engineering Design it is important to learn how to better integrate and take advantage of open data sources in research projects. An important driver for this position is the need to increase replicability, which improves our means to test the validity and reliability of research results of and within the field. As important as replicability is the possibility to build, not only on each other's final research results, but also on the used data; something that is more feasible when using open data sources. Although we have argued that open data presents an interesting opportunity for the Engineering Design field, there are also several limitations, risks, opportunities and challenges that are important to understand in order to better support the use of open data when such data is available.

5.1 Limitations and risks related to the use of open data

Unlike the case of closed data that is gathered for a specific research purpose, open data exists for the researcher as secondary data, often produced for entirely different purposes. This generates a lack of control in terms of data inputs, making information hard to validate without additional data gathering. In the context of Engineering Design research, a discipline characterised by a highly context-dependent design process, this means that results arrived at through the use of open data may need to be contextualised and triangulated with other data acquired specifically for the research purpose.

As with other data-driven research, researchers using open data also need to be aware of the limits and risks of generating theories based on automated or semi-automated analyses of large bodies of data. This includes the difficulties of distinguishing between “the signal and the noise” (Silver, 2012) and the risk of over reliance on complex algorithms and data harvesting techniques that are sometimes poorly understood. To mitigate these risks, researchers can use a combination of a robust theoretical background with empirical on-the-ground evaluations, to test if the patterns and predictions found in the models are good representations of real design phenomena.

5.2 Opportunities and challenges for research and practice

Beyond the previously stated limitations and risks, the inclusion of open data in Engineering Design research and practice also brings interesting opportunities and new challenges that can be used as a lever to further develop the discipline and its industrial impact.

For instance, the combination of big and open data, although technically challenging, presents an important opportunity to explore engineering design phenomena from new angles. These types of datasets are special, not only because of their openness, but also because they contain rich longitudinal data for a number of variables, at multiple scales, and from all Engineering Design domains, including product, organisation and increasingly also process data (e.g. open source design). Here "quantity has a quality of its own", and to take advantage of the volume, variety and velocity of data new methods are required, including advanced capabilities to join datasets from disparate sources.

5.3 Towards a stronger integration of open data in the research agenda

The challenge of integrating open data in Engineering Design research is systemic. It requires more resources for data infrastructures and capability building, improved policies, a revision of the importance of openness in our research cultures, better institutional incentives, and a good amount of creativity and perseverance to recognise where, among the large piles of open data available, might lay relevant and useful information. All this sounds like an exciting new “engineering design challenge” to be solved by our research community.

REFERENCES

Aitamurto, T., Holland, D. and Hussain, S. (2013), “Three layers of openness in design: Examining the open paradigm in design research”, *ICED 2013 - International Conference on Engineering Design*.

Albers, A., Ebel, B. and Sauter, C. (2010), “Combining Process Model and Semantic Wiki”, *Proceedings of DESIGN 2010, the 11th International Design Conference*, pp. 1275–1284.

Bartling, S. and Friesike, S. (2014), *Opening Science The Evolving Guide on How the Internet Is Changing Research, Collaboration and Scholarly Publishing*, *Opening Science*, available at:https://doi.org/http://dx.doi.org/10.1007/978-3-319-00026-8.
Bianchi, J., Knopper, Y., Ozgur, E., Badke-Schaub, P. and Roussos, L. (2015), "Online Ways of Sharedness: a Syntactic Analysis of Design Collaboration in Openideo", *Proceedings of the 20th International Conference on Engineering Design (ICED15)*.
Blessing, L. and Chakrabarti, A. (2009), *DRM, a Design Research Methodology*, Springer, London.
Cash, P., Stanković, T. and Štorga, M. (Eds.). (2016), *Experimental Design Research*, Springer International Publishing, Cham, available at:https://doi.org/10.1007/978-3-319-33781-4.
Chai, K.-H. and Xiao, X. (2012), "Understanding design research: A bibliometric analysis of Design Studies (1996–2010)", *Design Studies*, Elsevier Ltd, Vol. 33 No. 1, pp. 24–43.
Chiarini, M., Debra, T. and Hutchison, D. (2014), *Advancing the Impact of Design Science: Moving from Theory to Practice*, edited by Tremblay, M.C., VanderMeer, D., Rothenberger, M., Gupta, A. and Yoon, V., Vol. 8463, Springer International Publishing, Cham, available at:https://doi.org/10.1007/978-3-319-06701-8.
Fichter, K. (2009), "Innovation communities: the role of networks of promotors in Open Innovation", *R&D Management*, Vol. 39 No. 4, pp. 357–371.
Gillespie, D. (2010), *Engineering Visions in Design*.
Hamraz, B., Caldwell, N.H.M. and Clarkson, P.J. (2013), "A Holistic Categorization Framework for Literature on Engineering Change Management", *Systems Engineering*, Vol. 16 No. 4, pp. 473–505.
Hanson, B., Sugden, A. and Alberts, B. (2011), "DATA: Challenges and Opportunities", *Science*, Vol. 331 No. 6018, pp. 692–693.
Hey, T., Tansley, S. and Tolle, K. (Eds.). (2009), *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Microsoft Research, Redmond, Washington.
Hidalgo, C.A., Klinger, B., Barabási, A.L. and Hausmann, R. (2007), "The product space conditions the development of nations", *Science*, American Association for the Advancement of Science, Vol. 317 No. 5837, p. 482.
Kang, S.W. and Tucker, C. (2015), "An automated approach to quantifying functional interactions by mining large-scale product specification data", *Journal of Engineering Design*, Vol. 4828 No. January, pp. 1–24.
Le, Q. and Panchal, J. (2012), "Analysis of the interdependent co-evolution of product structures and community structures using dependency modelling techniques", *Journal of Engineering Design*, Vol. 23 No. 10–11, pp. 807–828.
Luo, J., Olechowski, A.L. and Magee, C.L. (2014), "Technology-based design and sustainable economic growth", *Technovation*, Elsevier, Vol. 34 No. 11, pp. 663–677.
MacCormack, A., Rusnak, J. and Baldwin, C.Y. (2006), "Exploring the Structure of Complex Software Designs: An Empirical Study of Open Source and Proprietary Code", *Management Science*, Vol. 52 No. 7, pp. 1015–1030.
Makumbe, P., Seering, W. and Rebentisch, E. (2009), "Beyond Cost: Product Complexity and the Global Product Development Location Advantage", *International Conference on Engineering Design, ICED ´09*.
Michelino, F., Lamberti, E., Cammarano, A. and Caputo, M. (2015), "Measuring Open Innovation in the Bio-Pharmaceutical Industry", *Creativity and Innovation Management*, Vol. 24 No. 1, pp. 4–28.
Nielsen, M. (2011), *Reinventing Discovery: The New Era of Networked Science*, Princeton University Press.
Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S.J., Buck, S., et al. (2015), "Promoting an open research culture", *Science*, Vol. 348 No. 6242, pp. 1422–1425.
OECD. (2015), *Data-Driven Innovation Big Data for Growth and Well-Being*, OECD Publishing.
Open Knowledge International. (2012), "What is Open Data?", available at: http://opendatahandbook.org/guide/en/what-is-open-data/ (accessed 1 November 2016).
Pajo, S., Vandevenne, D. and Duflou, J.R. (2015), "Systematic Online Lead User Identification - Case Study for Electrical Installations", *20th International Conference on Engineering Design (ICED 15)*.
Pampel, H. and Dallmeier-Tiessen, S. (2014), "Open Research Data: From Vision to Practice", *Opening Science*, Springer International Publishing, Cham, pp. 213–224.
Parraguez, P. and Maier, A.M. (2012a), "Towards describing co-design by the integration of Engineering Design and Technology and Innovation Management literature", in Hansen, P.K., Rasmusssen, J., Jørgensen, K. and Tollestrup, C. (Eds.), *NordDesign Conference 2012*, Center for Industrial Production, Aalborg University, Aalborg, pp. 465–481.
Parraguez, P. and Maier, A.M. (2012b), "Mapping industrial networks as an approach to identify inter-organisational collaborative potential in new product development", in Hernández-Cuevas, C., Hontavilla, J.P., Navarrete, M.J., Sandoval, V. and Garretón, M. (Eds.), *Encuentros Paris 2012 – Knowledge for Economic and Social Development*, Encuentros, Paris.
Parraguez, P. and Maier, A.M. (2016), "Network Insights for Partner Selection in Inter-organisational New Product Development Projects", in Dorian, M., Mario, S., Neven, P. and Nenad, B. (Eds.), *International Design Conference - Design 2016*, Dubrovnik, Croatia, pp. 1095–1104.

Pourmohamadi, M. and Dong, A. (2011), “The network topology of open innovation freelancing”, *ICED 2013 - International Conference on Engineering Design*.
Provost, F. and Fawcett, T. (2013), *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*, O’Reilly, Sebastopol.
Raghupathi, D., Yannou, B., Farel, R. and Poirson, E. (2014), “Sentiment rating algorithm of product online reviews”, *Proceedings of International Design Conference, DESIGN 2014*.
Silver, N. (2012), *The Signal and the Noise: Why so Many Predictions Fail--but Some Don’t*, Penguin Press, New York.
Song, B., Triulzi, G., Alstott, J., Yan, B. and Luo, J. (2016), “Overlay Patent Network for Analyzing Design Space Evolution: the Case of Hybrid Electrical Vehicles”, *DS 84: Proceedings of the DESIGN 2016 14th International Design Conference*, pp. 1145–1154.
Song, B., Yoon, B., Lee, C. and Park, Y. (2016), “Development of a service evolution map for service design through application of text mining to service documents”, *Research in Engineering Design*, Vol. In press, available at:https://doi.org/10.1007/s00163-016-0240-5.
Spee, J. and Basaiawmoit, R.V. (2016), “Design Thinking and the Hype Cycle in Management Education and in Engineering Education”, *Proceedings of the DESIGN 2016 14th International Design Conference*, pp. 2111–2124.
Staudenmayer, N., Tripsas, M. and Tucci, C.L. (2005), “Interfirm Modularity and Its Implications for Product Development”, *Journal of Product Innovation Management*, Vol. 22 No. 4, pp. 303–321.
Stone, T.M. and Choi, S. (2013), “Consumer Preference Estimation From Twitter Classification: Validation and Uncertainty Analysis”, *ICED13: 19th International Conference on Engineering Design*.
Storga, M., Mostashari, A. and Stankovic, T. (2013), “Visualisation of the organisation knowledge structure evolution”, *Journal of Knowledge Management*, Vol. 17 No. 5, pp. 724–740.
Tuarob, S. and Tucker, C.S. (2015), “Automated Discovery of Lead Users and Latent Product Features by Mining Large Scale Social Media Networks”, *Journal of Mechanical Design*, Vol. 137 No. 7, p. 71402.
Tucker, C. and Kim, H. (2011), “Predicting emerging product design trend by mining publicly available customer review data”, *18th International Conference on Engineering Design, ICED 11*.
Unger, D.W. and Eppinger, S.D. (2009), “Comparing product development processes and managing risk”, *International Journal of Product Development*, Vol. 8 No. 4, p. 382.
Walthall, C.J., Devanathan, S., Kisselburgh, L.G., Ramani, K., Hirleman, E.D. and Yang, M.C. (2011), “Evaluating Wikis as a Communicative Medium for Collaboration Within Colocated and Distributed Engineering Design Teams”, *Journal of Mechanical Design*, Vol. 133 No. 7, p. 71001.
Wang, M., Chen, W., Huang, Y., Contractor, N.S. and Fu, Y. (2016), “Modeling customer preferences using multidimensional network analysis in engineering design”, *Design Science*, Vol. 2 No. e11, available at:https://doi.org/10.1017/dsj.2016.11.
Wang, Z., Childs, P.R.N. and Jiang, P. (2013), “Using Web Crawler Technology To Support Design-Related Web Information Collection in Idea Generation”, *ICED13: 19th International Conference on Engineering Design*.
Zhou, Y.M. (2013), “Designing for Complexity: Using Divisions and Hierarchy to Manage Complex Tasks”, *Organization Science*, Vol. 24 No. 2, pp. 339–355.

ACKNOWLEDGMENTS

This research is partially funded by the Danish Industry Foundation (Industriens Fond) through the Net-Sights project (www.netsights.dk). The authors would also like to thank State of Green and CLEAN for their input and project support.